DOCUMENT RESUME

ED 370 680 PS 022 325

AUTHOR Sanogo, Youssouf; Gilman, David
TITLE Class Size and Student Achievement: Tennessee's Star and Indiana's Prime Time Projects.
PUB DATE Apr 94
NOTE 40p.
PUB TYPE Reports - Evaluative/Feasibility (142)

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS *Academic Achievement; *Class Size; Elementary School Students; Evaluation Problems; Longitudinal Studies; Primary Education; Program Evaluation; Research Design; *Research Methodology; *Small Classes; *Teacher Student Ratio
IDENTIFIERS Hawthorne Effect; Indiana; John Henry Effect; *PRIME TIME Project IN; Project STAR; *Student Teacher Achievement Ratio Project TN; Tennessee

ABSTRACT

This study investigated the possible causes of the contradiction between the results of two projects. Indiana's Prime Time project compared the achievement of students in large (averaging 26 students) and small (averaging 19 students) classrooms in grades 1 through 3. Results indicated that 3 years in smaller classes had little effect on student achievement. Tennessee's Student Teacher Achievement Ratio (STAR) project was a longitudinal study of class-size effects on student achievement in kindergarten through grade 3. The study concluded that small classes (13-17 students) had an advantage over large classes (22-26 students) in reading and mathematics. The present study examined whether students in the small classes in the STAR program really learned more than students in the large classes, and offered four hypotheses: (1) there was a relationship between the methodologies of the two projects and the contradictions in their results; (2) a Hawthorne effect occurred in the STAR program, according to which students in experimental groups tried harder than students in control groups; (3) a John Henry effect occurred in the STAR program, according to which students in control groups did not try harder than students in experimental groups; and (4) the research methodology of the STAR project was no better than that of the Prime Time project. The present study collected information about both projects' methodologies, designs, and circumstances. The study concluded that evidence did not definitively confirm a Hawthorne or John Henry effect, and that the STAR methodology was not better than the Prime Time methodology. (TM)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED 370 680

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
X This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

CLASS SIZE AND STUDENT ACHIEVEMENT: TENNESSEE'S STAR AND INDIANA'S PRIME TIME PROJECTS.

By
Youssouf Sanogo and David Gilman

INDIANA STATE UNIVERSITY

PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
David A. Gilman

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)"

April, 1994

PS 022325

BEST COPY AVAILABLE

I. ABSTRACT:

This study investigated the cause of the contradiction between the Tennessee's Student Teacher Achievement Ratio (STAR) project results and Indiana's Prime Time results. The methodologies and designs of both projects were checked, as well as the circumstances that brought them about. It was found that the methodologies and designs had a strong relationship with the observed contradiction. A type 1 error was found in the results reported in the STAR final executive summary. In other words, actually, the research which investigated Prime Time and STAR projects would have had similar results if the STAR research had been conducted in a less biased manner. The contradiction was explained by the fact that Tennessee Association of Education could influence the STAR project results, while the few evaluations of Prime Time were done by independent researchers and were not controlled by either the Indiana State Teachers Association or the Indiana Department of Education.. A strong probability of Hawthorne effect was also found in the STAR study.

II. BACKGROUND OF THE PROBLEM:

One of the questions in education that remains without any clear-cut answer is " Are smaller classes better ? " That question has been asked since 1900 (Swan, Stone, and Gilman, Aug. 1985), and research has never been able to answer it once and for all. So far, all the research studies about class size and student achievement have given results that , most of the time, tend to confuse more than they clarify about the issue.

Common sense would answer " smaller is better " for some reasons that Swan et al. (1985) categorized as:

"1. Teachers would have the energy and interest to give more concerned care and attention to each child if there are fewer in the classroom.

2. Classroom management is more effective when teachers spend more time with each student and keep track of individual progress.

3. Teachers will be able to employ a wider variety of instructional strategies, methods, and learning activities and can be more effective with them when class size is small.

4. Teachers' attitudes and morales are more positive when they have fewer students.

5. Small class size makes good use of added time and space.

6. Teachers will be able to find more time to plan, diversify, and individualize their teaching.

7. As teacher attention, energy, and time are shared among fewer students, the environment will be more conducive to learning."

Put that way, these reasons are likely to convince parents, school officials and policy makers that small classes are a sine qua non condition for a better student achievement. For example, in the last ten years, in Mali, elementary school teachers generally complained about class size when they talked about the low achievement of their students. They found the class size too large to be managed effectively. For instance, the teacher of a class of 60 students would find it extremely difficult to have enough time to pay attention to each and every student, because they were too many. The low achievement of elementary students was, most of the time, attributed to that situation, and teachers had arguments to defend their position. The complaints were still going on in 1992, and it was finally decided that the government and the communities should build more classrooms. So, billions of Malian currency had to be invested in construction, while officially no scientific research had been done to check if small classes improved student achievement.

Since the late 1970's, in USA, educators have made serious research studies about the relationship between class size and student achievement, but the results appear to be still confusing . In 1978, Glass and Smith conducted a massive literature review of essentially all 20th century research on class size and student achievement, and made a meta-analysis. They found that :

- there was a strong relationship between class size and student achievement.

- student achievement would rise by almost 1/2 standard deviation if classes were reduced to 15 students.

- achievement would rise by nearly 1 standard deviation if classes were reduced to about 5 students (Odden, Allan; 1990).

There were problems with Glass and Smith's study. Their analysis was based on the analysis of only 14 of the 77 studies they reviewed. They had chosen those 14 studies for their methodological soundness and dropped the remaining 63 for the simple reason that they were not true experimental studies. Another problem with Glass and Smith's meta-analysis was that some of the chosen studies that produced large effects were on learning how to play tennis (Odden, 1990).

Glass and Smith's findings were strongly criticized. The Educational Resources Service, Inc. (E.R.S) declared that research findings on class size and student achievement were inconclusive and contradictory (Mulder,1990). Thus, a debate that lasted a decade took place between Glass and E.R.S. Slavin (1990), another critic of the Glass and Smith's study, he reviewed Glass and Smith's study and concluded that " learning benefits do not appear until class size is reduced to three. " According to him, dramatic achievement effects can be obtained from one-to-one tutoring.

Other researchers, Gilman (1993) and Harder (1990), found that instructional effectiveness depends more on the teacher and the quality of instruction than on class size. Gilman pointed out that " class sizes in schools have been going down since the 1920's, and test scores have often been going down along with them. " He suggested that rather than reduce class size, it would be more appropriate to attack school discipline problems directly and find a way to

help students who have behavior problems or deal with them in a way that they will not keep students who want to learn from learning. " As for Harder, she affirmed that " class size should not become a smoke screen to draw attention away from the real issue, which is the quality of education. "

In any case, it is obvious that reducing class size will require more classrooms, teachers and supplies. For instance, as estimated by the U.S. Office of Education (Gilman, 1993), the reduction of every public school class size would require 33 % increase in educational costs, including: 73.3 billion dollars per year in teachers' salaries, 47 billion dollars per year in indirect costs (fringe benefits, furniture, instructional materials, building expenses) and an additional hiring of 1,365,821 teachers. Tomlinson (1989) found that the policy of reducing class size was not only impractical, but also, far from raising the quality of classroom instruction, it might well lower it. To him, least qualified local teachers would have a much greater chance to be hired, and that would do less for children's education.

Yet, despite the cost and the controversy around class size and student achievement, most teachers and parents prefer smaller classes. Tomlinson (1989) said that at least 18 states intended to adopt the small class policy. Nowadays, it has become necessary more than ever to determine , once and for all, whether smaller classes are better to avoid the adoption of the wrong solution to the wrong problem: spending huge amounts of money for class reduction while the problem of student achievement is not there.

Recently, both Indiana and Tennessee have made scientific research to

check if students achieve better in smaller classes, in early elementary schools. Educators really expected much from both studies. The first was Indiana's project Prime Time (1984-87). It was a large scale study supported by the Indiana Department of Education. Prime Time compared scores in reading, mathematics, writing, and composite of large classes to the ones of reduced classes in grades 1,2, and 3. The larger classes averaged 26.9 students per class and had experienced no Prime Time classes. The smaller classes averaged 19.1 students per class and had experienced Prime Time classes in grades 1, 2, and 3. The results indicated that three years in a reduced class size had little effect on the academic achievement of primary students (Gilman & Tillitsky, 1989).

At the same time, a similar project was engaged in Tennessee, Tennessee's project Student Teacher Achievement Ratio (STAR) (1985-89). It was a longitudinal study of class- size effects on pupil achievement and development in early primary grades (K-3). The research was based on reading and mathematics. It compared achievement scores of small classes (13-17 students per teacher) to regular classes (22-26), and regular classes with full time teacher aide (22-26) (Achilles, Bain, and Finn; 1991). The conclusion of the final executive summary report stated,

" This research leaves no doubt that small classes have an advantage over larger classes in reading and mathematics in the early grades. This experiment yields an unambiguous answer to the question of the existence

of a class-size effect, as well as estimates of the magnitude of the effect for early primary grades " (Word, Achilles, Bain, Folger, Johnston, and Lintz; 1990)

The conclusion of the STAR project contradicted the project Prime Time conclusion, and probably created confusion in the minds of many educators. Yet, both research projects were large scale studies and were conducted by professionals in the same time frame. If Prime Time found that smaller classes had little effect on early elementary students achievement whereas STAR found the contrary, then the cause of the contradiction must have been the methodologies and designs.

III. STATEMENT OF THE PROBLEM:

The results of the research studies about the relationship between class size and student achievement have not really determined whether smaller classes result in greater achievement. The research results are contradictory and controversial. Despite that situation, with the generally increasing desire of parents and teachers for small class reduction, the small class policy is likely to be (if not is) a fashion in education. The general question behind this study was " Does class size have any impact on early elementary students achievement ? " More specifically, this study investigated if early elementary students really learned more in smaller classes in Tennessee's STAR project. Four hypotheses were investigated:

1. There is a strong relationship between the methodologies and designs of Tennessee's STAR and Indiana's Prime Time, and the contradiction

between the results of the two projects.

2. The experimental group students and teachers in Tennessee's STAR project knew they were in an experimental group and tried harder to get a better achievement than the students and teachers in the control group. In other words, there was a Hawthorne effect.

3. The control group students and teachers of Tennessee's STAR knew they were in a control group and did not try to get better performance than the experimental group students and teachers. In other words, there was a the Hawthorne Effect was operating here.

4. The research methodology and design of Tennessee's STAR were no better than the research methodology and design of Indiana's Prime Time.

METHODOLOGY:

Data about the class size issue were articles and reports pertaining to research studies conducted about class size and student achievement from 1978 to 1990, including Indiana's Prime Time and Tennessee's STAR.

The articles and reports were read to collect data about the background of the class size issue, Prime Time and STAR projects. The example about Mali was a testimony. The main focus of the study was on Tennessee's STAR and Indiana's Prime Time projects: methodologies and designs. However, important information was also found about circumstances that brought them about.

1. Tennessee's STAR project: It was a four year study (1985-89) to get a definitive answer to the question of the effects of class size. It was provided 3

million dollars per year to implement the research design. Four universities (Memphis State University, Tennessee State University, University of Tennessee, Knoxville, and Vanderbilt University) provided technical assistance in the design and the conduct of the study. More guidance about a number of design characteristics was also given by Tennessee's legislation.

STAR project was conducted in inner city, suburban, urban, and rural schools; in the east, middle and west Tennessee. The class types were: small classes (13-17 students per teacher), regular classes (22-25), and regular with full time-teacher aide (22-25). Small classes and regular classes with full time-teacher aide were the experimental group, and the control group was the regular classes. The study covered kindergarten and grades 1,2,and 3. Student achievement was the primary criterion for judging the effectiveness of the class size reduction. Student development was also measured (Folger, March 1989; Word, E. et al; June 1990).

2. Indiana's Prime Time project:

Prime Time (1984-87) was also a state-wide project. Indiana had passed legislation to spend $150-180 million to fully implement it (Malloy and Gilman, May 1988). Prime Time was supported by the Indiana Department of Education. It studied the effect of Prime Time on 52 schools and 30 school districts and made comparisons between scores in reading, mathematics, writing, and composite subtest scores of large classes with the ones of reduced classes in grades 1,2, and 3. The larger classes averaged 29.9 students per class, and the smaller ones averaged 19.1 students per class. The larger classes had no Prime Time

experience while the smaller classes experienced Prime Time classes for three years. The objective of the study was to determine the effects of class size on student achievement (Gilman & Tillitsky , 1989). Small classes were the experimental group while large classes were the control group.

Data collected about the circumstances, the methodologies and designs were examined, discussed and interpreted to test hypotheses, draw conclusions, and make recommendations. No further computation was made.

V. RESULTS:

The data collected about the circumstances, the methodologies and designs were summarized:

1. Tennessee's STAR: circumstance, methodology and design:

STAR grew out of a controversy about Governor's Alexander's Better Schools program in 1983. The centerpiece of that program was the master teacher program which would evaluate teachers and pay better teachers more. The Better Schools program was strongly opposed by the Tennessee Education Association. An alternative was finally reached and consisted of lowering class size in the early elementary grades from the existing maximum of 25 to 21 per class. The cost of the alternative was equal to the cost of the master teacher program (about 80 to 100 million dollars per year).

At first, the Governor and the legislature opposed the alternative, but Representative Steve Cobb, chief sponsor of the Better Schools program in the House, was interested in the class size issue. He decided that the effects on student achievement of a class size reduction in grades k-3 to 15 students per

class should be demonstrated. Representative Steve Cobb had reviewed Glass's meta-analysis, and been told about preliminary results of class size study in one Nashville school by Helen Bain and Charles Achilles. He expected the STAR project to be a definitive study that would establish for Tennessee and other states with similar early elementary school programs the size of the class size effect (Folger, Fall 1989).

Dr. Bain was the one who had urged the legislature to fund a statewide class size study. He was a strong advocate for reduced classes. She and Achilles were among those who wrote the final executive summary of STAR project (Folger, 1989; Word et al., 1990).

Four universities , (Memphis State University, Tennessee State University, University of Tennessee, Knoxville; and Vanderbilt University) contracted with Tennessee's State Department of Education to design, study, collect, analyze the data, and develop the final report of the project. An external advisory committee was also set up.

The districts that participated in the project were not randomly selected since the participating schools in each district were volunteers. Project schools had average tests scores slightly below the state-wide average, because there was a higher proportion of inner-city schools with low test scores in the sample than in the whole state. Their class size was above the state average class size (.4 of a pupil in the year before the project began). They were also 6% above the state average in per-pupil expenditures and 2% above the state average teacher salaries. Despite these differences, the project staff concluded that the

sample schools were representative of all schools in Tennessee (Folger, 1989).

A " within school" design was made to reduce major sources of variations in student achievement attributable to school effects. Each school was required to have at least 57 students at the appropriate grade level so that it could contain at least one of each class type (small, regular, and regular with aide). In each year of the study, there were more than 6,000 students. The number of Subjects varied for several reasons, including that kindergarten was not required in Tennessee

(Word, E., et al; June 1990).

A three day in-service training was organized in thirteen (13) schools to train teachers to optimize their instructional effectiveness. Fifty seven (57) teachers got special training in the second grade and fifty five (55) in the third grade. Some teachers didn't get any special training. In each school, teachers were observed once teaching reading and mathematics lessons to help them optimize their instructional effectiveness, including non-trained teachers (Folger, Fall 1989).

Each year the teachers were randomly assigned to one of the three class types by the project staff. Initially, students were randomly assigned to a class type and they stayed with that class type throughout the project. The new students were also assigned randomly to class type in accordance with vacancies. By the project fourth year, about one-third of the students had been in the same class type all four years, and the other two-thirds were replacing and added students (Folger, Fall 1989).

The project final executive summary (Word et al. 1990) stated that:

- The student achievement was measured by the appropriate forms of the Stanford Achievement Test (k-3), the STAR's Basic Skills Criterion Tests (grades 1-2), and Tennessee's Basic Skills Criterion (grade 3). Student development was measured by the Self-Concept and Motivation Inventory (SCAMIN).

- The results showed a definite advantage for students in small classes in achievement and no significant advantage for the use of teacher aide. Small classes students outperformed students in regular and in regular with aide classes by substantial (statistically and educationally significant) margins on standardized test and on the Basic Criterion Tests of reading and mathematics. This pattern continued in grades 2 and 3 as shown in figures 1 and 2.

_ Figures 1 and 2 here_

-In the third grade total reading and total mathematics scaled scores and percentile ranks by location and class type, the greatest advantage was for inner-city small classes. The highest scores in all class types were made in rural schools. The least advantage was for regular with aide classes in urban and suburban schools. Longitudinal results for the small (about 33%) subsample of students in the same class size for two (k-1) and three years (1-3) showed that the large statistically significant gains favoring the small classes made in the first year (i.e. K in the K-1 comparison, and grade 1 in the 1-3 comparison) were maintained as shown in figures 3 and 4.

_ Figures 3 and 4 here _

The Self-Concept and Motivation Inventory (SCAMIN) revealed that students in small classes in kindergarten had significantly higher self concept score. Being in a small class did not have any impact on student self-concept or motivation in grades 1 through 3. Statistically significant findings based upon school location snowed that inner-city (predominantly minority) students had higher self-concept scores in grades 1 and 2, and they also had higher motivation scores in grade 3.

However, another study (Folger and Breda, 1989) showed that surveys of project STAR teachers indicated that almost all of them believed that smaller classes were better. Two-thirds of the teachers said they would prefer a one-third smaller class to a $ 2,500 a year raise. Another study (Odden, 1990) informed that in a recent solid longitudinal study (Folger, 1990) almost no achievement differential was found for STAR third grade students who had been in smaller classes since kindergarten.

Prime time was proposed by Robert D. Orr, Governor of Indiana, and Harold H. Negley, former Superintendent of Public Instruction (Varble and Gilman, 1988). The pilot study started in 1981 and lasted two years. It took place in twenty four (24) kindergarten through second grade classes in nine (9) schools across Indiana and reduced the student/teacher ratio to 14:1. It was reported to be successful after two semesters as students exceeded normal achievement in both reading and mathematics. As a result of that success, Prime Time was conducted in all first grade classes in Indiana in 1984-85.

However, Gilman, Swan, and Stone (1988) concluded that the pilot study

was conducted by teachers carefully chosen rather than being selected through traditional hiring or assignment practices. In addition, although many variables were measured in the study, only those that produced significant results were reported.

In the same study (Gilman et al., 1988), it appeared that the Department of Education officials were reluctant to conduct a state-wide study to evaluate the results of the program. The only attempts to evaluate Prime Time were subjective observations of the activities in carefully selected school systems by six evaluators who were carefully controlled by the Department of Education Staff. Moreover, it should be noted that (Gilman, 1993) " the policy of the Indiana Department of Education (indeed its first policy statement) has been that they only conduct and fund research that supports the policies of the Indiana Board of Education. "

Prime Time was not implemented on a uniform basis:

- A few teachers received inservice training in small class teaching strategies while most did not.

-In some schools, teachers were given large classes (over 24) and provided with aides instead of having class size reduction. Some aides were trained and others were not.

-In some small communities Prime Time did not reduce class size.

-In most school systems, there was no formal evaluation of Prime Time. However, in some school systems, teachers were told that gains in student achievement were expected. In some cases, teachers were informed of evaluative studies to

be conducted at the end of the year, and in other cases the evaluation was unannounced (Gilman and Antes, 1985).

Diverse tests were administered during Prime Time. The Iowa Test of Basic Skills (ITBS), the Stanford Achievement Test (SAT), the California Achievement Test (CAT), and the Indiana Competency Test (ICT) were frequently used (Gilman and Tillistky, June 1989).

The ITBS results of Prime Time three-year cohort study (Tillitsky, Gilman, Mohr, and Stone, 1988) in the North Gibson School Corporation in Princeton, Indiana, showed that gains favoring small classes that were evident in grades 1 and 2 had largely disappeared by the end of grade 3.

_ Table 1 and figure 5 here_

A longitudinal study (Gilman, and Tillitsky, 1989) examined the effect size of Indiana's Prime Time on student achievement in Southwestern Indiana. According to the 1980 US Census Information characteristics of race, education, and income in Southwestern Indiana are comparable to the state demographics. The effect size for each test was predicted by Wolf's weighted mean method. Wolf's average effect size method was used to predict the average for all schools. Seventy six (76) comparisons of achievement test results were made for twenty seven (27) selected schools. Scores for a total of 2,333 students were analyzed for the larger class and were compared to a total of 2,272 students in smaller class group. The results showed that:

- Of the 26 comparisons for the Reading Subtest, 14 favored the smaller Prime Time classes and 12 favored the larger classes.

- Of the 26 comparisons for the Mathematics Subtest 14 favored the smaller Prime Time classes and 12 favored the larger classes.

- In Writing Subtest scores, only 1 of the 5 comparisons favored smaller classes and 4 favored the larger classes.

- In Composite Subtest scores, Prime Time classes were favored in 10 of the 20 comparisons, and 10 favored larger classes.

For the total of all comparisons, 39 favored the Prime Time group and 38 favored the larger classes.

The statistics for the total effect size, when all of the comparisons were combined, showed that the effect size was 0.02 standard deviation units for reading, -0.01 for mathematics, -0.13 for writing, and 0.001 for composite.

The total for all comparisons was 0.01 standard deviation units. So three years in a reduced class size environment had got little effect on students academic achievement.

_ Table II here _

VI. DISCUSSION, CONCLUSIONS AND RECOMMENDATIONS:

As expected for the first point investigated, there was a strong relationship between the methodologies and designs, and a contradiction was observed between the two results. But, unexpectedly, it was also found that the circumstances that brought the projects about strongly influenced their implementations. The contradiction was mostly due to the fact that Indiana Department of Education didn't make a state-wide evaluation whereas Tennessee Association of Education did. In fact, the few evaluations of Prime

Time made in some schools in Indiana were not carefully controlled by the State Department of Education staff, whereas Tennessee Association of Education could influence the STAR project results. If the evaluators of Indiana Department of Education had evaluated Prime Time, there would be a different story: Prime Time results might have been similar to the results of STAR.

Moreover the study published by Folger (1990) revealed that the gains of small classes in kindergarten and grades 1 and 2 had almost disappeared in grade 3. Apparently, there was no contradiction between the results of Prime Time and the results of STAR. Three or four years in small classes had no significant effect on student achievement.

The motivation behind the type 1 error of STAR was to convince the Governor, and the Tennessee Legislature to drop the Better Schools Program and adopt the small class policy. The same motivation was behind the amazing results of inner city and rural students. Obviously there was politics in Tennessee STAR.

For the second point investigated, absolute affirmation could not be made. However, a strong probability of Hawthorne effect existed. The districts that were chosen and the school systems that volunteered to be in the sample were aware of the challenge of the project. Officials and teachers of those districts and school systems believed in the small class policy and knew that positive experimental results might lead to a policy of small classes. Since the first beneficiaries of the small class were the teachers, experimental group teachers (particularly those in small classes) would try harder so that their students could

perform better than those in the control group. With such an attitude there was more chance that experimental group students be made aware of being in experimental group, at least to increase their motivation to some extent. Yet, the surprise might be the total disappearance of gains in score of the regular with aide classes (another experimental group) in the third grade. However, that situation might have been caused by other factors related to the novelty of the approach and a problem of compatibility of some teachers and their aides.

The third point investigated revealed that the same attitude behind the Hawthorne effect, would affect the behavior of the control group teachers and students, and cause a John Henry effect. The control group teachers, knowing the benefits of small classes for their profession, would not worry about trying hard with the students to get a better performance. That idea coupled with their own belief in the small class policy gave a strong probability of a John Henry effect. But no evidence was found that would allow to say that the control group students were made aware of being in control group and tried no harder to perform better than experimental group students.

As for the fourth point investigated, the expectation was met. Technically, Tennessee's STAR project methodology and design were no better than the research methodology and design of Prime Time, because both were not implemented with any scientific attitude. In fact, the research to study the STAR project was more elaborate than the research which investigated the Prime Time project. Technical assistance in the design and the conduct of the study was provided by a four-university consortium, a "within-school" design was made,

students were formally evaluated; but the whole effort became worthless because the sample of the study was biased.

As for Indiana's Prime Time, it was conducted in all first grade classes, and lacked a uniform implementation, a formal evaluation in most school systems, and a state-wide evaluation. Only the results of some schools were studied and evaluated by some researchers (different from the project evaluators). In any case, the Indiana Board of Education was biased since the beginning of Prime Time, and failed to evaluate the project seriously. The pilot study was biased. It was conducted in a way that showed the Board of Education's intention to introduce a small class policy would be successful. The Department of Education officials' reluctance for a state-wide evaluation of Prime Time.

However, the difference in elaboration between Prime Time and STAR projects was not enough to state that STAR was better than Prime Time in terms of methodology and design, because technically, neither were true experimental research.

Although a strong probability of Hawthorne and John Henry effects existed, further studies are still needed to determine for sure:

1. Whether STAR experimental group teachers did make their students work harder for a better performance.

2. Whether STAR control group teachers did not try hard to get better performance from their students.

For the future, it's important that more attention should be given to not only the elaboration of the methodology and design of research studies, but also

to their implementation. In addition, for the reliability and the validity of the studies more effort needs to be made to avoid bias in sampling. Policy makers of education should also overcome their own emotions and adopt a scientific attitude for the benefit of schools.

As for the relationship between class size and student achievement, it would be wiser to observe and think over the factors involved in a learning environment in classrooms. It would be good to pose the questions in terms of relationship between instructional techniques, curriculum, and student achievement. Students and teachers, like any other human beings, have emotions and other psycho-social characteristics. A low or high achievement of students cannot be explained solely by the number of students in a class.

REFERENCES:

1. Achilles, C. M.; Bain, H. P.; and Finn, J. D. (March, 1991). Why Won't Class Size Issues Just Go Away ? Tennessee's Student Teacher Achievement Ratio (STAR) Project (8/85-8/89). Paper presented at the National Meeting of the American Association of School Administrators, Orlando, FL : 5.

2. Folger, John. (March, 1989). Project STAR : Design and Measurement. EDRS, No. ED 307 697: 3-4.

3. — (Fall, 1989). Project STAR and Class Size Policy. *Peabody Journal of Education*, 67. 1: 7-12.

4. Folger, John; and Breda, Carolyn. (Fall, 1989). Evidence From Project STAR About Class Size and Student Achievement. *Peabody Journal of Education*, 67. 1:17.

5. Gilman, David Alan. (1993). "Small Classes : Miracle or Myth." *Terre Haute Tribune -Star*, 24 February : 1.

6. Gilman, David Alan ; and Tillitsky, Christopher. (June, 1989). The Longitudinal Effect Size of Prime Time, Indiana's State Sponsored Reduced Class Size Program. Indiana State University : 2-5, 9-11.

7. Gilman, David Alan ; Swan, Edward ; and Stone William. (Winter,1988). The Educational Effects of a State Supported Reduced Class Size Program : A Comprehensive Evaluation of Indiana's ProjectPrime Time at the North GibsonSchool Corporation. *Contemporary Education*, 59. 2 : 113

8. Gilman, David A, and Antes, Richard, (Fall, 1985). The Educational Effects of the Introduction of a State-Supported of Smaller Classes. A Study the First Yearof Indiana's Project Prime Time and Its Effects on Test Results (1984-85). A Comprehensive Analysis. EDRS, No. ED 286 842 : 6-7.

9. Harder, Heather. (Fall, 1990). A Critical Look at Reduced Class Size. *Contemporary Education*, 62. 1 : 29.

10. Malloy, Leane; Gilman, David. (May, 1988). The Cumulative Effects of Indiana Prime Time : A State -Sponsored Reduced Class Size Program, on Basic Skills Achievement. Indiana State University, EDRS, No. ED 300 139 : 6.

11. Mulder, Jordan. (Fall, 1990). Class Size Revisited : The Glass/ERS Debate. *Contemporary Education*, 62. 1 :47-49.

12. Odden, Allan. (Summer, 1990). Class Size and Student Achievement : Research-Based Policy Alternatives.Educational Evaluation and Policy Analysis, 12. 2 : 215-16, 218.

13. Slavin, Robert. (Fall,1990). Class Size and Student Achievement : Is Smaller Better ? *Contemporary Education*, 62. 1 : 7.

14. Swan, Edward; Stone, William; and Gilman, David Alan. (Aug. 1985). The Educational Effects of a State Supported Reduced Class Size Program. A Comprehensive Evaluation of Indiana's Project Prime Time at the North Gibson School Corporation. Terre Haute : Indiana State University, School of Education. EDRS, No. ED 276 109 : 4-5.

15. Tillitsky, C.; Gilman, D.; Mohr, A.; and Stone, W. (Fall, 1988). Class Size Reduction in North Gibson School Corporation : A Three-year Cohort Study. ERS Spectrum, 6. 4 : 38-40.

16. Tomlinson, Tommy M. (1989). Class Size and Public Policy : Politics and Panaceas. Educational Policy, 3. 3 : 262.

17. Varble, Mary Ellen, and Gilman, David Alan. (March, 1988). A Study of the Relationship Between Class Size and Achievement. Indiana State

BEST COPY AVAILABLE

University, EDRS, No. ED 291 527 : 4.

18. Word, Elizabeth; Achilles, Charles M.; Bain, Helen; Folger, John; Johnston, John; and Lintz, Nan. (June, 1990). Project STAR Final Executive Summary Report, Kindergarten Through Third Grade (1985-1989). North Nashville, TN 37219 : Tennessee State Department of Education, Project STAR Office, Cordell Hull Building, Fourth Floor.

BEST COPY AVAILABLE

APPENDIX

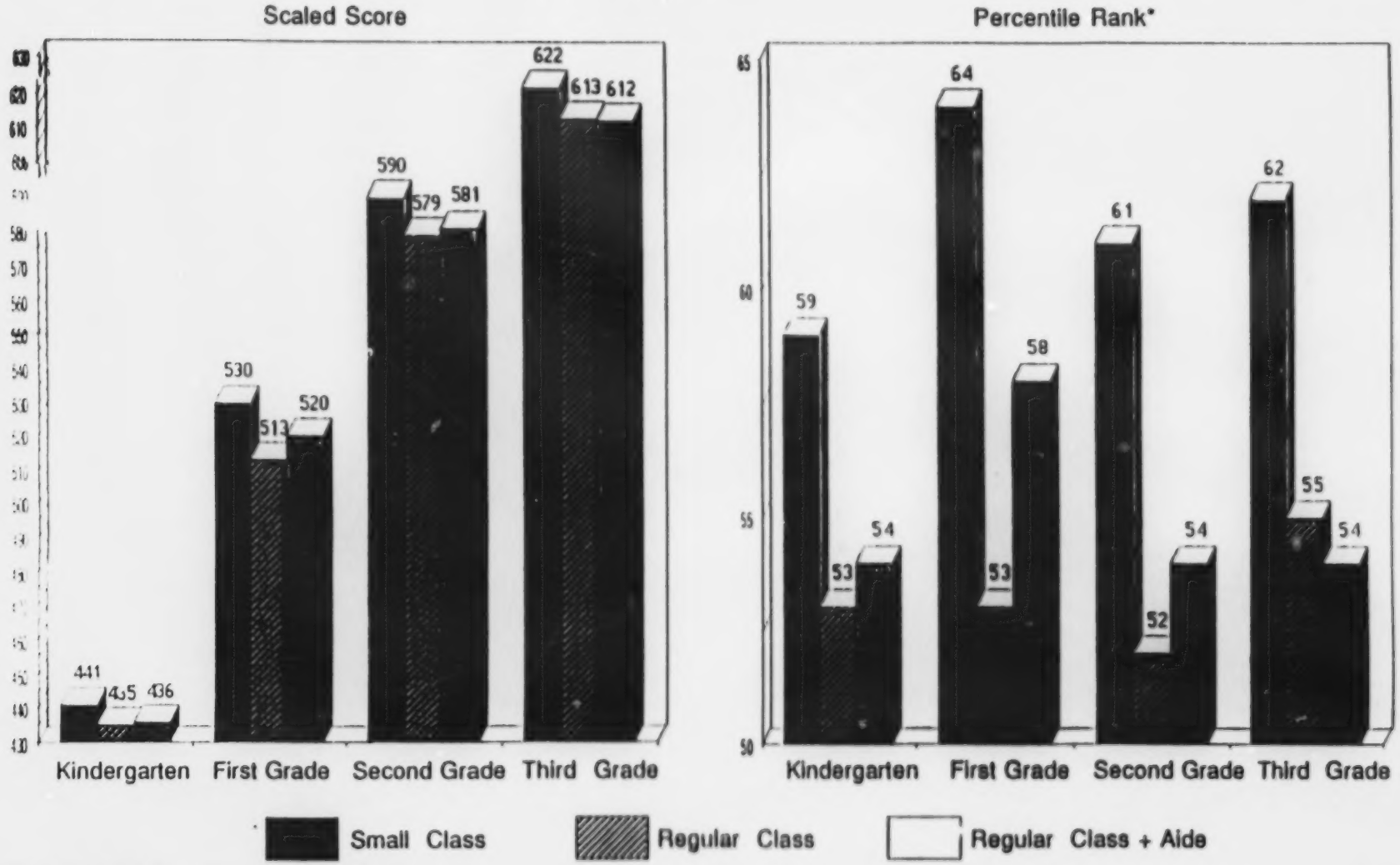
APPENDIX

Figure 1

Project STAR

Stanford Achievement Test

Total Reading: Class Type by Grade



Stanford SESAT II, Primary I, II, and III

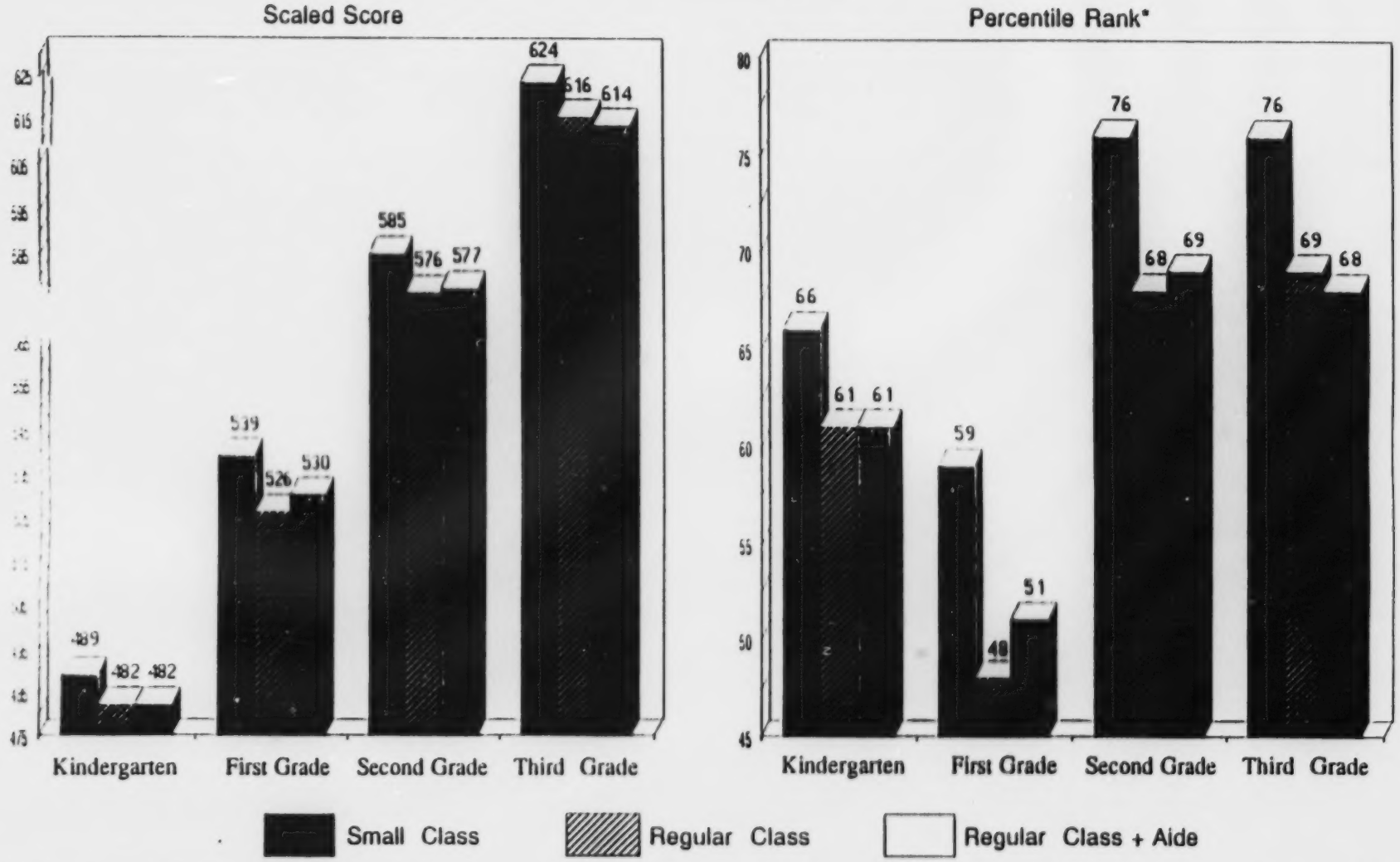
*Percentile rank is based on Stanford Multilevel Norms

Figure 2

Project STAR

Stanford Achievement Test

Total Math: Class Type by Grade



* Stanford SESAT II, Primary I, II, and III

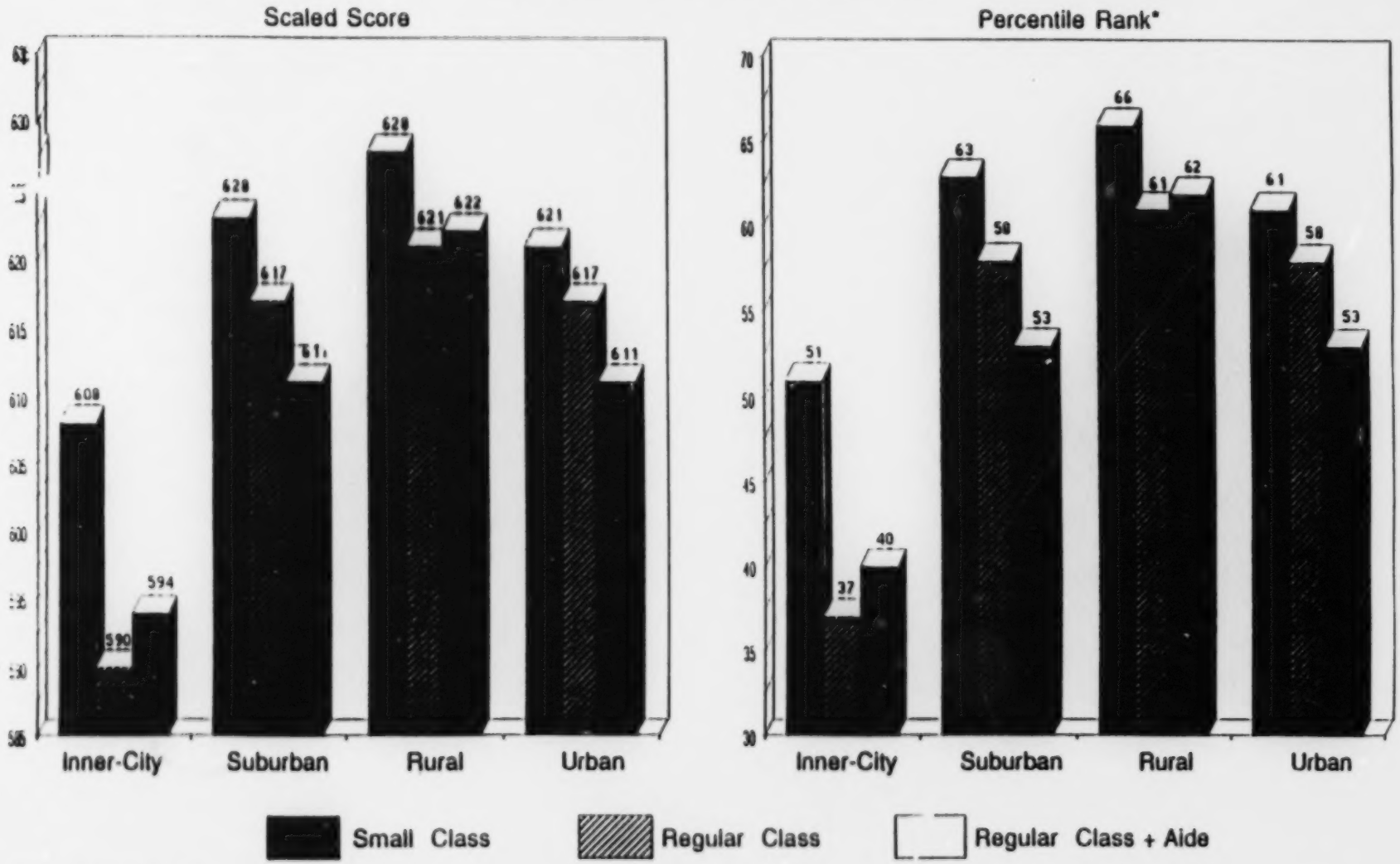
*Percentile rank is based on Stanford Multilevel Norms

Figure 3

Project STAR

Third Grade Stanford Achievement Test

Total Reading: Class Type by School Type



Stanford Primary III

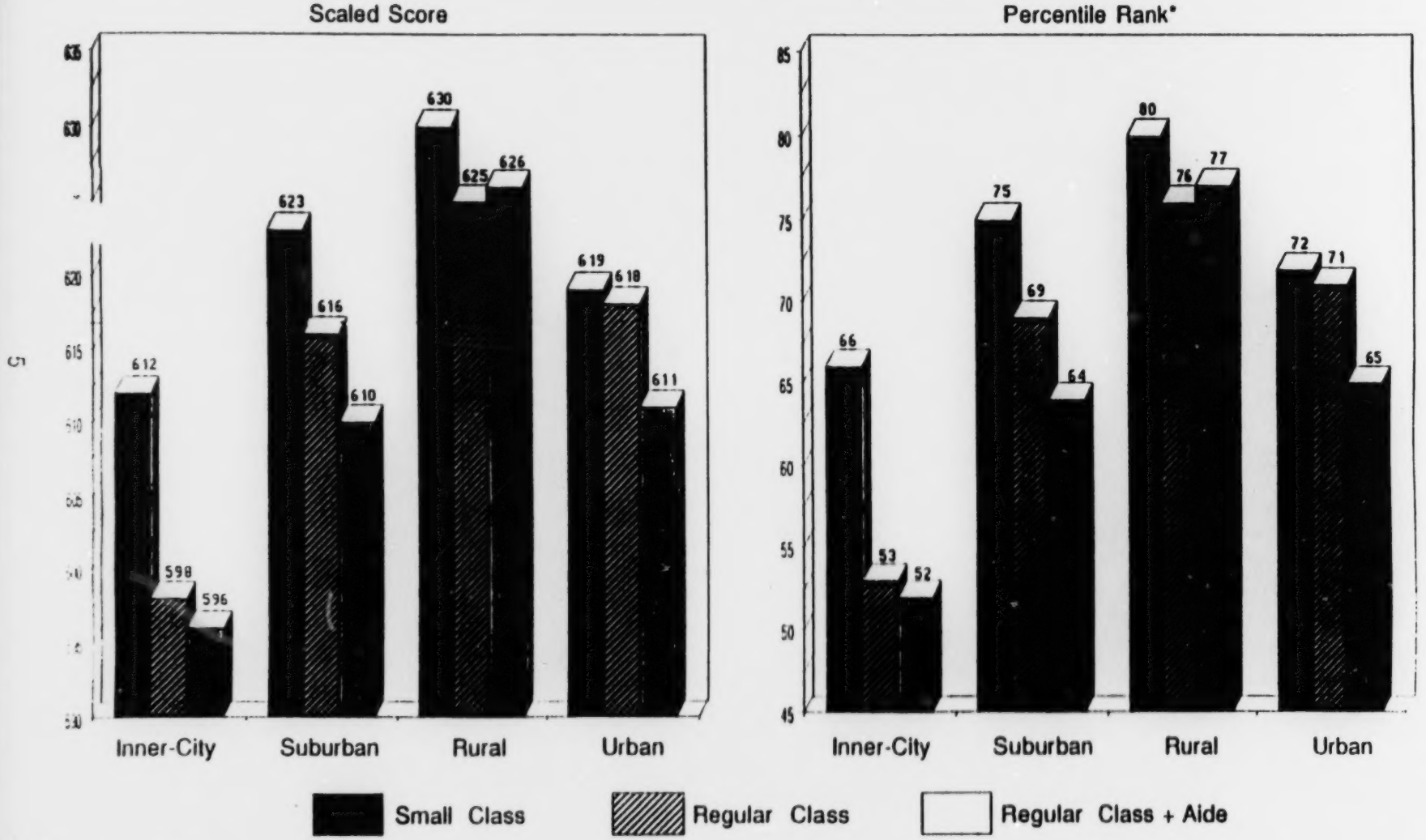
*Percentile rank is based on Stanford Multilevel Norms

Figure 4

Project STAR

Third Grade Stanford Achievement Test

Total Math: Class Type by School Type



Stanford Primary III

*Percentile rank is based on Stanford Multilevel Norms

Table 1.— Results on Iowa Test of Basic Skills for Large Class Cohort and PRIME TIME Cohort, First Grade Through Third Grade

READING SUBTEST

	First Grade			Second Grade			Third Grade		
	mean	sd	size	mean	sd	size	mean	sd	size
Large Class	71.5	19.9	23.7	68.9	21.2	20.5	66.0	20.9	24.0
PRIME TIME	75.2	17.8	19.9	72.4	22.5	17.4	64.6	23.6	18.0
F ratio		6.04			1.22			-0.22	
p<		.001			.27			.64	

MATH SUBTEST

	First Grade			Second Grade			Third Grade		
	mean	sd	size	mean	sd	size	mean	sd	size
Large Class	66.2	23.7	23.7	58.3	26.1	20.5	71.1	23.1	24.0
PRIME TIME	76.5	22.3	19.9	71.6	24.1	17.4	74.9	21.1	18.0
F ratio		9.54			13.67			1.38	
p<		.002			.003			.24	

COMPOSITE SUBTEST

	First Grade			Second Grade			Third Grade		
	mean	sd	size	mean	sd	size	mean	sd	size
Large Class	75.0	16.7	23.7	68.6	22.4	20.5	71.9	19.5	24.0
PRIME TIME	79.9	17.1	19.9	77.5	19.6	17.4	72.2	20.4	18.0
F ratio		4.12			8.51			0.01	
p<		.04			.004			.91	

Figure 5.—Results on Iowa Test of Basic Skills, Large Class Cohort and PRIME TIME Cohort, Grades One-Three